



LABORatorio R. Revelli
Centre for Employment Studies

Firm Data Analysis in Linked Employer-Employees Datasets

Roberto Leombruni

Agosto 2003

Firm Data Analysis in Linked Employer-Employees Datasets

Roberto Leombruni – LABORatorio R. Revelli

leombruni@labor-torino.it

Abstract

In linked employer-employee databases, when the employees are sampled first and then the employers data are attached, the set of firms that enter the sample is biased towards the large dimension. In this paper I discuss two strategies to impute inflating factors to the employers' records, and to obtain a sample representative of the entire population of firms. A comparison between the two strategies is made with the help of Monte Carlo simulations. As an application, I build a rotating panel of Italian firms from a linked employer-employee longitudinal database of administrative source for the years 1986 to 1998, and compare some stylised facts derived on it with the existing knowledge on Italian firms' size distribution.

Keywords: Firm Microdata; Linked Employer-Employees Datasets; Firms' Size Distribution

Jel codes: C13; C81; J63; L1

Introduction

When building a linked employer-employee database, the sample designer has two opposite options at her choice. She can sample the workers first, and then link into the database the records of the firms where they are employed; or she can sample the firms first, and then link into the database all (or a sample of) their employees.

There are substantive and economic reasons for using either of the two strategies. As regards substantive reasons, the unit that is sampled first (from now on, the sampled unit) is likely to be the primary longitudinal unit, and the sample designer has to choose it bearing in mind which is the main focus of the database. As regards economic reasons, when the data source is an original survey to be implemented, there is a certain bias towards the choice of sampling the firm first, since in this way – with a given costs budget – higher sample dimensions can be achieved (Krebs et al., 1998). When the data source is administrative, on the other side, economic reasons are less noticeable, since typically the switch among the two strategies will just involve some editing in the piece of code that queries an already existing database.

The choice of the sampled unit has a crucial side effect that should be kept into consideration: The units that are matched to the sampled unit (from now on, the linked units) will usually not provide us with a sample representative of their target population. Indeed, when some of the characteristics of the linked unit have an influence on the probability that they are “attached” to the sampled ones, the set consisting of the units linked will have a selection bias.

Consider the case in which a random sample of firms is drawn, and then all the individuals there employed are matched into the database. Let us assume that older workers have a higher probability than young ones to be working in larger firms (as they usually have), and that the firms are drawn with a constant sampling probability. The employees of large firms will then be over represented, and the resulting workers’ sample will be biased towards older workers.

In general, while the representativeness is granted for the sampled unit, to make inference about the population of the linked one a bit of work is to be done to calculate their probability of being in the sample, and to impute then a correct inflating factor.

In this paper I analyse the case of employer-employee longitudinal datasets where the sampled unit is the worker. I derive two strategies to impute inflating factors to the employers’ records, and to obtain this way two panels representative of the entire population of firms (section 1). Pros and cons of the two strategies are discussed with the help of Monte Carlo simulations (section 2). In the final section I apply the procedure to the Panel LRR, an employer-employee longitudinal dataset of administrative source covering Italian dependent employment, and compare some stylised facts coming out from the sample with the existing knowledge on Italian firms’ size distribution as available from other firms data sources.

The firm sample

Our task in this section is the following: Starting from a longitudinal linked employer-employee dataset, where the sampled unit is the worker, build a probability sample of firms.

To put it formally, let us call W the target population of individuals w , and F the target population of firms f . The job relations between workers and firms can be summed up with a dynamic index function J_t defined on the Cartesian product $(W \times F)$, such that $J_t(w, f) = 1$ when at time t the worker w is employed in firm f , and equal to zero otherwise.

Let us call $T = \{1, 2, \dots, t, \dots, \tau\}$ the set of τ time units covered by the longitudinal dataset. We will consider the following subsets of workers:

$W_{t,f} = \{w \in W : J_t(w, f) = 1\}$ is the set of individuals employed in firm f at time t ;
 $W_{T,f} = \bigcup_{t \in T} W_{t,f}$ is the set of individuals that have been employed in firm f at any time during T ;
 $W_t = \bigcup_{f \in F} W_{t,f}$ is the set of individuals employed (in any firm) at time t ;
 $W_T = \bigcup_{t \in T} W_t$ is the set of individuals that have been employed at any time during T .

With this notation, we can restate our task as follows: Given a couple $P_T = \{W \subseteq W_T, F(W) \subseteq F\}$, where W is a set of workers sampled with a sampling probability π , and F is the set of firms linked to the workers sampled, find out the probability for a generic firm to enter in the set F .

For a firm to enter in the sample it is sufficient that one or more of its employees have been extracted. The event “Firm f entered in the sample F ” hence is equivalent to the event “Any of the workers that have been employed by firm f during T has entered the sample W ”. To compute the corresponding probability then we have to span with a logic *or* the set $W_{T,f}$:

$$\Pr\{f \in F\} = \Pr\{\bigcup_{w \in W_{T,f}} (w \in W)\} \quad (1)$$

If the workers are sampled randomly, the probability for an individual to be extracted is independent from what happens to other workers. The events $w' \notin W$ and $w'' \notin W$ ($w' \neq w''$) then are independent, and the probability (1) can be rewritten as

$$\begin{aligned} \Pr\{f \in F\} &= \Pr\{\bigcup_{w \in W_{T,f}} w \in W\} \\ &= 1 - \Pr\{\bigcap_{w \in W_{T,f}} w \notin W\} \\ &= 1 - (1 - \pi)^{\#(W_{T,f})}. \end{aligned} \quad (1')$$

where $\#(W_{T,f})$ is the cardinality of $W_{T,f}$ — that is, the number of workers transited in the firm during T .

When we observed $\#(W_{T,f})$, then, we could use formula (1') to weight each firm $f \in F$ with the exact *ex ante* probability of entering in the sample, producing this way a probability sample representative of the target population of firms. The fact is, that we usually do not observe it.

The information that we usually have is the sequence of firm sizes at each $t \in T$. Because of the turnover, however, the set of workers employed by a firm at a given $t' \in T$ can be very different from the set of workers it is employing at a different $t'' \in T$, and the number of workers transited will then be higher than the average size of the firm during T .

Two strategies can be employed to overcome the difficulty. The first one is to try and estimate the number of workers transited. Assuming for the sake of simplicity that workers hired by a firm have not had other job relations with the same firm during T , the number of worker transited will be the following:

$$\#(W_{T,f}) = S_{f,1} + H_{f,1} + H_{f,2} + \dots + H_{f,t} + \dots + H_{f,\tau} \quad (2)$$

where $S_{f,1}$ is the size of the firm in the initial period, and $H_{f,t}$ is the number of hires in time t . If we are not far from a situation of equilibrium, we can approximate (2) with the following:

$$\#(W_{T,f}) \approx S_{f,\cdot} (1 + \tau h_{f,\cdot}) \quad (2')$$

where $S_{f,\cdot}$ is the average firm size in T , and $h_{f,\cdot}$ is the average hiring rate in T , $< H_{f,t} / S_{f,t} >$.

Till now, we have just expressed the number of workers transited in terms of the hiring rate and the length of the panel. Our problem is not solved yet, since we cannot observe the hiring rate at the firm level. Job relations, however, can be observed also from the worker side of the data, so that hires *are* observed. This means that we can have a correct estimate of an average hiring rate simply computing the so called entry rate e , defined as the ratio of job accessions in a time unit on total employment.

If we plug this average value into (2'), and substitute into (1'), we obtain a formula for the probability for a firm of being extracted, as a function of its average size, the length of the panel, and an estimate of the turnover rate of its workforce:

$$\Pr\{f \in F\} = 1 - (1 - \pi) S_f(1 + \tau e). \quad (1'')$$

Note that a firm with a higher than average level of turnover will have a higher *ex ante* probability to enter in the sample with respect to other firms of the same size class. Since we are unable to tell in our sample high from low turnover firms, a selection bias will remain, since – as we may assume – the turnover level is probably correlated with many of the firm's characteristics on which we are willing to make inference. In order to investigate the firm size distribution, however, the formula (1'') provides us with correct estimates, since it corrects exactly for the bias introduced by the turnover into the size distribution. We will get back to this point in the next section.

If the frequency at which firms' data are collected is sufficiently high we have at our disposal a second strategy, that at the cost of some loss of information provides us with a more effective firm sample. The strategy is: For each t , select only those firms that in the same t are currently employing a worker of the sample, and impute a weight according to formula (1'), using the firm size instead of $\#(W_{T,f})$. If, in a subsequent t' , there are no more of its employees in the sample, the firm will exit from the rotating panel.

The crucial assumption is that data have to be observable at a high frequency. If this is the case, within each time t the turnover is negligible, the number of workers transited is equal to the firm size, and formula (1') gives us the correct inflating factors. The impact of turnover on a longer time scale, on the other side, is ruled out by the fact that we are selecting for each t just the firms having an active job contract with some workers present in the sample.

A drawback of this methodology is that we lose some information: To make inference in a given t , in fact, we are discarding all those firms that are present in the set F , are possibly active at time t , but with no of their time t -employees in the worker sample. When appropriate, the efficiency loss can be reduced pooling the information over T .

Monte Carlo simulation

In this section, I get deeper into the pros and cons of the proposed strategies by means of a Monte Carlo microsimulation.

I considered a family of target populations spanning different values for the turnover level. For each target population I extracted a battery of linked employer-employee samples (LEE), sampling the workers first and then linking the employers. For each LEE extracted, I produced the firm samples according to the two strategies proposed, and computed the pivot statistics with which to compare them.

The target populations are built around a set of firms each represented by a vector $f_{it} = \{S_{it}, S^*_{it}, h_{it}, f_{it}\}$, where S and S^* are the actual and the desired number of employees respectively, h is the probability to fill an open position in the time unit, and f is the probability of firing an employee in the time unit. Each of the four variables are heterogeneous in the population: S^* is extracted from an exponential distribution, while the hire and fire rates are extracted from uniform ones. The

population of employees is generated filling each of the firms' positions with a generic worker – for our purposes, no worker characteristics are needed to be considered either than a unique ID. At time $t = 0$, then, actual and desired size are equal for each firm.

To evolve the initial population, I implemented for each simulated time unit the following schedule of events:

1. Each firm fires some of its employees;
2. All the individuals without a job are randomised into an unemployed queue;
3. Each firm tries and fills some of its opened positions, calling individuals from the unemployed queue;

With these assumptions, the firm actual size changes over time according to the following:

$$\begin{aligned} E[S_{t+1}] &= S_t + E[H_{t+1}] - E[F_{t+1}] \\ E[F_{t+1}] &= S_t f \\ E[H_{t+1}] &= [S^*_t - S_t (1 - f)] h \end{aligned}$$

where H and F are hiring and firings, and the index i has been dropped for ease of readability. The process has a stable equilibrium defined by the condition $E[H_t] = E[F_t]$. In equilibrium, then, the expected ratio between open vacancies and desired size for each firm is equal to:

$$E[(S^* - S)/S^*] = f(1 - h) / [h + f(1 - h)]$$

Since the sum of desired firm sizes is equal in our simulation to the number of workers, this ratio is equal to the unemployment rate, and has been used as a control value for the outcomes of the simulation.

Within this framework, I considered the following parameters setting. I built the initial populations with 5.000 firms with an average size (and standard error) of 40 employees, for an average of 200.000 workers for each target population simulated. Since the level of turnover is the crucial variable for us, the family of target populations created spans different values for the hiring and firing probabilities. For each population, firing and hiring probabilities have been uniformly extracted from the intervals $[0, \alpha]$ and $[0, 2\alpha]$ respectively, with α going from 2.5% to 20% (step 2.5). This way, I had a family of target populations covering both very low and very high turnover levels. Each initial population has been evolved for 30 periods – all simulations started fluctuating around the unemployment equilibrium value as soon as after 10-15 time units.

For each population created, then, I extracted a battery of a hundred LEEs, sampling the workers with a probability of 1%, and then linking in the panel their work histories and their employers for the time interval $T = \{21, 22, \dots, 30\}$. I then considered the following firm samples:

1. An unweighted sample F_u , composed by all linked firms “as are”, without any selection or inflating factor imputation;
2. An entry rate weighted sample F_{ew} . The same as F_u , plus an inflating factor that weights the firms according to formula (1”) – i.e., approximating the number of workers actually transited by means of an average entry rate;

3. A size weighted rotating panel F_r . A rotating panel composed in each $t \in T$ by those and only those firms with an active job contract with a worker sampled, weighted using the formula (1'') and assuming turnover to be negligible within the time unit ($e_f = 0$)¹.

To have a sketch of the selection bias of the unweighted sample F_u , in figure 1 I compare the actual size distribution in one of the target population created, with an instance of a corresponding unweighted sample F_u . In the unweighted sample the size distribution is quite shifted towards big dimensions, since small firms have a probability to enter in the unweighted sample much smaller than big ones.

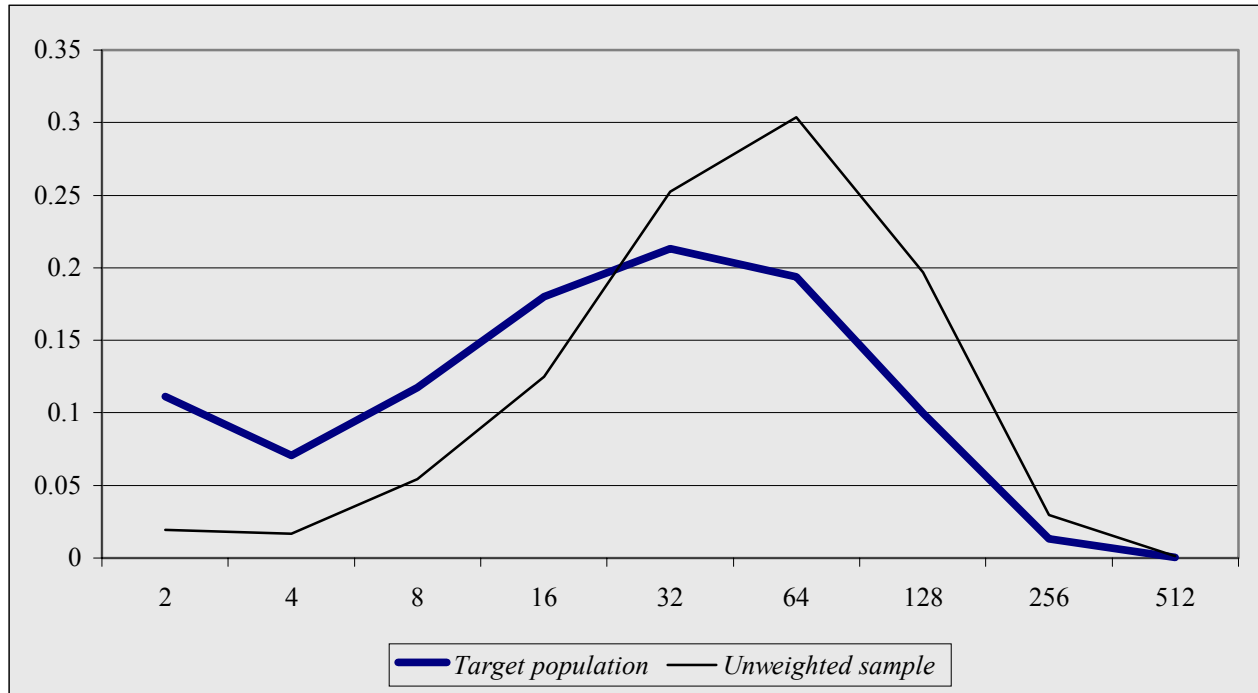


Fig. 1. Firms' size distribution in the target population and in a linked unweighted sample.

Note: Parameters' values as in the text, with $\alpha = 20$.

The comparison between the two strategies considered is the object of tables 1 and 2. Table 1 reports the average values and standard deviations of the average firms sizes, computed on the hundred samples created, and compares them with the true value in the corresponding target population. Panel (A) is relative to the hypothesis of low turnover level. Both probability samples give estimates quite close to the true value of 27.792. As expected, the loss of information for the size weighted rotating panel is reflected by a higher value for the standard deviation. The indications that we can derive from panels (B) and (C), relative to target populations with higher level of turnover, go in the same direction.

A bit surprising, on the contrary, is the fact that the standard deviation for the entry rate weighted samples gets lower. In fact, the sampling probabilities are estimated ones, by means of the sample entry rates. On the other side, higher turnover levels in the simulations are associated with higher heterogeneity – remember that hire and fire probabilities were extracted from intervals $[0, \alpha)$ getting more and more wide. One should expect then that increasing the turnover level the estimates for the entry rates got more and more uncertain, and this should result in estimates for the firm size more uncertain too. There is a countervailing factor, however: The higher is the average turnover

¹ To fulfill the last requirement – unrealistic with the parameters values hypothesised – I simulated the firing and hiring events as happening exactly between the time units, and selecting firms according to the workers employed strictly within the time units.

level, the larger are the probabilities for every firm to be extracted. The resulting firm sample will then be larger, and the rise in the sampling ratio countervails the efficiency loss that can be produced by more uncertain estimates of the entry rate.

To make inference about the size distribution of firms, then, both strategies proposed in section one appear to be good ones, and if one can rely on good estimates for the entry rates, the sample F_{ew} probably gives more efficient estimates.

**Table 1. Comparison between the two sampling strategies.
Estimates of the average firm size.**

(A) Low turnover level ($\alpha = 2.5\%$)		
	average	standard deviation
Value in the target population:	27.792	
Entry rate weighted sample	27.706	0.935
Size weighted rotating panel	28.308	1.087
(B) Medium turnover level ($\alpha = 10\%$)		
	average	standard deviation
Value in the target population:	26.769	
Entry rate weighted sample	25.867	0.869
Size weighted rotating panel	27.759	1.225
(C) High turnover level ($\alpha = 20\%$)		
	average	standard deviation
Value in the target population:	28.412	
Entry rate weighted sample	27.005	0.822
Size weighted rotating panel	29.439	1.194

As anticipated, if we want to make inference about some other firm characteristics, the selection bias will probably get back. Even though the sampling probabilities estimated by means of the entry rate provide us with correct estimates for the size distribution, we nonetheless are still unable to tell high from low turnover firms. Moreover, we know that high turnover firms are over represented. So, if we are investigating some firm characteristic correlated with the turnover level, to work with the entry rate weighted sample will give us biased estimates.

As an example, I considered in table 2 a variable strictly connected with the turnover level, the firing probability. The panel (A) refers again to a situation of low turnover. The selection bias here does not produce sensible distortions, and we have indications similar to what we observed in table 1: Both strategies provide good estimates, the sample F_{ew} looks more efficient.

Increasing the turnover level the situation changes dramatically. As expected (the fire probability is positively correlated with the turnover level) the entry rate weighted sample largely overestimates the fire probability, of about 13% for $\alpha = 10\%$, and of over 17% with $\alpha = 20\%$.

The rotating panel F_r , on the other side, still gives estimates that are close to the true value, which is largely within the 99% confidence interval.

**Table 2. Comparison between the two sampling strategies.
Estimates of the average firing probability (values x100).**

(A) Low turnover level ($\alpha = 2.5\%$)		
	average	standard deviation
Value in the target population:	2.505	
Entry rate weighted sample	2.625	0.068
Size weighted rotating panel	2.502	0.075
(B) Medium turnover level ($\alpha = 10\%$)		
	average	standard deviation
Value in the target population:	10.230	
Entry rate weighted sample	11.587	0.224
Size weighted rotating panel	10.104	0.279
(C) High turnover level ($\alpha = 20\%$)		
	average	standard deviation
Value in the target population:	19.926	
Entry rate weighted sample	23.420	0.338
Size weighted rotating panel	19.738	0.671

A rotating panel of Italian Firms

As an application of the methodology proposed, I build in this section a rotating panel of Italian firms covering the years from 1986 to 1998. I extract it from the panel LRR, a linked employer-employee longitudinal database of administrative source, and compare some of the stylised facts derivable from it with the existing knowledge on Italian firms' size distribution.

The panel LRR sampling scheme is the following. A sample of 1/90 workers is extracted from the individual file of the Social Security Administration. This amounts to setting as a target population the generality of workers that have had in the period covered a job relation with a private firm as a dependent employee. Each year, the number of workers that enter in the sample is about 120-130 thousands. For each employee, then, all the records relative to the job spells experienced are extracted. Finally, the business register is matched to the job spell archive, so that we observe the life path over the period covered of all the firms where the sampled workers transited.

Over all the period, the matched firms are about 85.000 per year, to be compared to a target population of roughly 1.150.000 firms. When pooled together into an unweighted sample F_u , we have a set of about 120.000 records, where small firms are largely under represented with respect to large ones. With the sampling scheme above sketched, the probability to enter in the sample for companies above 500 employees is close to 100%, while for 10 employees firms the same probability can be estimated as ranging between 15 and 20%.

To resample the firms' records into a sample representative of the target population, the formula (1') is not directly applicable, since we do not have any information about the number of workers transited in each firm during the period. Potentially, we could try and estimate how large is turnover at the firm level in the case of large firms, since the bigger is the firm, the higher is the number of its employees that are sampled, and on which we could make inference. However, the firms for which we could obtain this measure are the same about which we do not need it, since their sampling probability is already known to be close to 100%. For small firms, on the other side, we do not know which is the number of workers transited and we can not estimate it.

We could try and build either of the two weighted samples F_{ew} and F_r . Remember that the entry rate weighted sample can be useful in making inference only about the size distribution of firms. In the case we are considering, this strategy is a bit uncertain, since typically the entry rates are highly variable across firms and time, and idiosyncratic factors have a prominent role in their explanation (see Haltiwanger 1996, Contini, 2002). The uncertainty in the entry rates estimates ends up in a less reliable imputation of the inflating factors, and the efficiency lead with respect to a rotating panel presumably gets smaller.

On the other side, we have all the necessary elements to build a rotating panel F_r . The job data have a monthly frequency, so that, if we select for each year just the subset of firms with an active job contract in a given month, the turnover effect along the year is ruled out. Moreover, the number of employees recorded for each month in the business register is not an average size, but the number of workers present in the payroll of that month, whichever the number of days they actually worked. In other words, we know exactly the number of workers transited in the month, and the probability for each firm to be matched in the sample in a given month can be calculated without approximations. To be true, we have the opposite problem that lead us to write equation (2): Given the number of workers transited in a firm in a month, find out which is its average size. The average monthly turnover rate, however, has a magnitude such that we may expect an over estimation of average size of about 1 to 2%. Actually, the official statistics published on the basis of the administrative source ignore this turnover bias.

In table 3, I compare the firm size distribution in the target population, and the size distribution that comes out from the rotating panel. The data about the target population (panel a) are those published in aggregated tables in the “Osservatorio delle Imprese” by the Social Security Administration (http://www.inps.it/doc/sas_stat/imprese/imprese.html). The rotating panel with which I calculated the size distribution of panel (b) has been built choosing December as the month in which stock sample the active jobs, since the size published in the “Osservatorio delle Imprese” refers to the same month. Apart from a slight shift from the medium size classes to the 1-9 one, the two distributions are very close each other.

Table 3. Comparison between firm size distribution in the target population and in the rotating panel. Years 1993-1995, percentage values.

(a)	<i>Inps data on entire population</i>					
	<i>size class</i>					
	1-9	10-19	20-49	50-199	200-499	500+
1993	85.46	8.47	4.21	1.54	0.22	0.10
1994	85.35	8.54	4.24	1.55	0.22	0.10
1995	85.23	8.58	4.29	1.57	0.22	0.10

(b)	<i>Estimates with the rotating panel</i>					
	<i>size class</i>					
	1-9	10-19	20-49	50-199	200-499	500+
1993	85.99	8.21	4.02	1.44	0.23	0.11
1994	86.12	8.15	3.96	1.43	0.23	0.10
1995	85.79	8.29	4.09	1.49	0.23	0.10

Source: Inps, “Osservatorio delle Imprese”, and our elaborations on Panel LRR.

Recently, a thorough information about Italian firms’ demography and size distribution has been produced within the Oecd’s *Firm level project*, and made available in aggregated tables (see Bartelsman et al., 2003; Ahn 2001). The primary source of data is represented also in this case by the administrative archives of the Social Security Administration. The period covered is wider with respect to Inps’ “Osservatorio delle Imprese”, but more outdated, going from 1986 to 1994.

In table 4 there a comparison of the average firm size resulting from the Oecd data and as estimated in the rotating panel. When compared to Oecd data, in the rotating panel there seems to be an overall underestimation of the average firm size (8.7 as compared to 10.4 in Oecd data). Actually, the discrepancy is between Oecd data and what is published by Inps’ Osservatorio. If we take the overall number of workers as published in the Osservatorio and divide it by the number of firms

with at least one employee, we obtain a total average firm size of 8.68, much closer to our data (cp. again http://www.inps.it/doc/sas_stat/impresse/impresse.html).

If we try to take into account the difference in the overall average, and compare the distribution of the average size across sectors, they look more similar, apart from some big differences for the sectors Electricity, gas and water supply; Finance, insurance, real estate and business services; and Community, social and personal services. These differences can be explained with some still open questions with the alignment in the classification used, since in Inps source archives firms are classified according to an Italian release of the standard Nace 1970 (Ateco81), that is not fully compatible with Isic Rev. 3. Actually, there is not an official conversion file mapping Ateco81 into Isic Rev. 3, and some macroscopic differences – such as the average size in the Finance sector – have probably to be imputed to the (still provisional) conversion tables that I’m using.

Table 4. Comparison between average firm size from Oecd data and from the rotating panel, by Isic Rev. 3 sectors, years 1993.

	<i>Oecd</i>	<i>F_r</i>
Agriculture, hunting, forestry and fishing	6.0	5.6
Mining and quarrying	15.4	11.9
Food products, beverages and tobacco	10.3	9.1
Textiles, textile products, leather and footwear	13.7	12.0
Wood and products of wood and cork	6.5	5.9
Pulp, paper, paper products, printing and publishing	13.7	13.2
Chemical, rubber, plastics and fuel products	29.7	28.3
Other non-metallic mineral products	15.8	14.8
Basic metals, metal products, machinery and equipment excluding transport	15.4	13.8
Transport equipment	109.6	85.3
Manufacturing, n.e.c; recycling	11.1	10.2
Electricity, gas and water supply	241.7	74.1
Construction	7.0	6.0
Business sector services	7.0	6.2
Wholesale and retail trade; hotels and restaurants	5.3	4.8
Transport, storage and communications	22.6	18.1
Finance, insurance, real estate and business services	8.5	30.8
Community, social and personal services	12.0	5.7
Total	10.4	8.71

Note: Oecd data are from Bartelsman et al., 2003.

Concluding, the longitudinal sample of Italian firms produced according to the methodology proposed in section 1, apart from some issues “under construction”, connected with sectoral classification standards, shows a high compatibility with the existing knowledge on the firm size distribution of firms. With respects both to the tables published by the Social Security Administration, and to the tables made available by Oecd’s *Firm Level Project*, the rotating panel produced offers a greater detail (the micro data of firms are available, covering the entire length of the panel), a longer time coverage (from 1986 to 1998, compared to 1994-1998 and 1986-1994 respectively) and the linkage with workers data.

References

Ahn, S. (2001), “Firm dynamics and Productivity Growth: A Review of Micro Evidence from OECD Countries”, *OECD Economics Department Working Paper* No. 297, Paris.

Bartelsman, E., S. Scarpetta and F. Schivardi (2003), “Comparative Analysis of Firm Demographics and Survival: Micro-Level Evidence for the Oecd Countries”, *OECD Economics Department Working Paper* No. 348, Paris.

Contini, B. (eds.) (2002), “Labour Mobility and Wage Dynamics in Italy”, Torino, Rosenberg & Sellier.

Haltiwanger, J. (1997), “Measuring and analyzing aggregate fluctuations: the importance of building from microeconomic evidence”, *Federal Reserve Bank of St. Louis Economic Review*, (May): 55–77.

Krebs H., Z. Patak, G. Picot and T. Wannell (1998), “The Development and Use of a Canadian Linked Employer-Employee Survey”, Statistics Canada, mimeo